

DEALING WITH MOMENT MEASURES VIA ENTROPY AND OPTIMAL TRANSPORT

FILIPPO SANTAMBROGIO

ABSTRACT. A recent paper by Cordero-Erausquin and Klartag provides a characterization of the measures μ on \mathbb{R}^d which can be expressed as the moment measures of suitable convex functions u , i.e. are of the form $(\nabla u)_\# e^{-u}$ for $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ and finds the corresponding u by a variational method in the class of convex functions. Here we propose a purely optimal-transport-based method to retrieve the same result. The variational problem becomes the minimization of an entropy and a transport cost among densities ρ and the optimizer ρ turns out to be e^{-u} . This requires to develop some estimates and some semicontinuity results for the corresponding functionals which are natural in optimal transport. The notion of displacement convexity plays a crucial role in the characterization and uniqueness of the minimizers.

1. INTRODUCTION

We consider in this paper the notion of *moment measure* of a convex function, which comes from functional analysis and convex geometry. Given a convex function $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$, we define its moment measure as

$$\mu := (\nabla u)_\# \rho, \quad \text{where } d\rho = e^{-u(x)} dx.$$

The connection of this notion with the theory of optimal transport is straightforward from the fact that, by Brenier's Theorem, the map ∇u will be the optimal transport map for the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$ from ρ to μ .

In a recent paper, Cordero-Erausquin and Klartag ([9]) studied the conditions for a measure μ to be the moment measure of a convex function. First, they identified that an extra requirement has to be imposed to the function u in order the problem to be meaningful. The main difficulty arises in case u is infinite out of a proper convex set $K \subset \mathbb{R}^d$. In this case one needs to require some continuity properties of u on ∂K . Without this condition, every measure with finite first moment can be the moment of a function u , which is in general discontinuous on $\partial\{u < +\infty\}$. Also, there is a strong non-uniqueness of u . On the contrary, if one restricts to convex functions u that are continuous \mathcal{H}^{d-1} -a.e. on $\partial\{u < +\infty\}$ (those functions are called *essentially continuous*), then there is a clear characterization: a measure μ is a moment measure if and only if it has finite first moment, its barycenter is 0, and it is not supported on a hyperplane. Moreover, the function u is uniquely determined by μ up to space translations.

In [9], the authors first prove that these conditions on μ are necessary, due to summability properties of log-concave densities, and they prove that they are sufficient to build u as the solution of a certain minimization problem.

Here we want to reprove the same existence result with a different method, replacing functional inequalities techniques with ideas from optimal transport. This aspect seems to be absent from [9] even if it is not difficult to translate most of the ideas and techniques of Cordero-Erausquin and Klartag into their optimal transport counterparts. The result is an alternative language, that

is likely to be appreciated by people knowing optimal transport theory, while the community of functional inequalities could legitimately prefer the original one. The question of which approach will the colleagues working with both optimal transport and functional inequality prefer is an open and unpredictable issue. . .

The main idea justifying this approach is the following: many variational problems of the form

$$\min \left\{ \frac{1}{2} W_2^2(\rho, \mu) + \int f(\rho(x)) \, dx \right\}$$

have been studied in recent years, for different purposes (time-discretization of gradient flows, urba planning. . . see for instance [7, 10, 12] and Chapter 7 in [13]). The optimality condition of this problem reads, roughly speaking, as

$$\phi + f'(\rho) = \text{const},$$

where ϕ is the Kantorovich potential in transport from ρ to μ for the cost $c(x, y) = \frac{1}{2}|x - y|^2$ (notice that the convex function appearing in Brenier's Theorem is given by $u(x) = \frac{1}{2}x^2 - \phi(x)$: if $T = \nabla u$ is the optimal transport, $-\nabla \phi$ is the optimal displacement, i.e. $T(x) = x - \nabla \phi(x)$).

In the case $f(t) = t \ln t$, the condition above implies that ρ is proportional to $e^{-\phi}$. In order to obtain the condition which is required we need to correct the above minimization, so that we get u instead of ϕ . In order to do so, one needs to change the sign and insert a $\frac{1}{2}|x|^2$ term, which leads to

$$\min \left\{ -\frac{1}{2} W_2^2(\rho, \mu) + \frac{1}{2} \int |x|^2 \, d\rho(x) + \mathcal{E}(\rho) \right\},$$

where $\mathcal{E}(\rho)$ denotes the entropy of ρ , defined as $\mathcal{E}(\rho) := \int \rho \ln \rho \, dx$ for $\rho \ll \mathcal{L}^d$, $\mathcal{E}(\rho) = +\infty$ for ρ non absolutely continuous. When μ has finite second moment and we minimize the above functional among measures ρ supported on a given compact set K it is easy to check that a minimizer exist and that we have $\rho = e^{-u}$, where $(\nabla u)_\# \rho = \mu$. Yet, it is highly possible that u is discontinuous at the boundary of K . Indeed, at least in the case where also μ is compactly supported, one sees that the function u must be Lipschitz (since its gradient is bounded), which means that it is bounded, and hence ρ is bounded from below by a positive constant. Yet, $\rho = 0$ and $u = +\infty$ outside K and u is not essentially continuous. This is a confirmation that every $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ is a moment measure, if we accept convex functions u which are not essentially continous.

The interesting case is the one where we minimize among measures $\rho \in \mathcal{P}_1(\mathbb{R}^d)$, without restricting their support to a compact domain. In this case the existence of an optimal ρ is not evident (indeed, the term $W_2(\rho, \mu)$ is continuous for the weak convergence of probability measures on compact sets, but only l.s.c. on unbounded sets, and here it is accompanied by the negative sign, which makes it u.s.c., while we want to minimize). Also, the lower semicontinuity of the entropy term $\mathcal{E}(\rho)$ is more delicate on unbounded sets. Here comes into play the assumptions on μ , which will allow to provide a bound on the first moment of a minimizing sequence ρ_n . Then, we can prove that ρ is log-concave and that a precise representative of ρ must vanish \mathcal{H}^{d-1} -a.e. on the boundary of its support, which proves that u is essentially continous. In order to handle the case of measures $\rho, \mu \notin \mathcal{P}_2(\mathbb{R}^d)$, we will profit of the fact that the second moment part of $\frac{1}{2} W_2(\rho, \mu)$ and $\frac{1}{2} \int |x|^2 \, d\rho(x)$ cancel each other, and that everything is well-defined for $\rho, \mu \in \mathcal{P}_1(\mathbb{R}^d)$ if we transform the minimal transport problem for the cost $c(x, y) = \frac{1}{2}|x - y|^2$ into the maximal transport problem for $c(x, y) = x \cdot y$.

Hence, we study the minimization problem

$$(P) \quad \min \left\{ \mathcal{E}(\rho) + \mathcal{T}(\rho, \mu) : \rho \in \mathcal{P}_1(\mathbb{R}^d) \right\},$$

where

$$\mathcal{T}(\rho, \mu) = \sup \left\{ \int (x \cdot y) d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}.$$

As we announced above, we prove that this problem has a log-concave solution of the form $\rho = e^{-u}$, that u is essentially continuous and μ is the moment measure of u . We also prove uniqueness up to translations of the minimizer and that the condition $\mu = (\nabla u)_\# \rho$, where $d\rho = e^{-u}$, is sufficient to minimize. This characterizes u . Uniqueness and sufficient conditions will be based on the notion of displacement convexity in the space $\mathcal{P}_2(\mathbb{R}^d)$ endowed with the distance W_2 . It is useful to notice that the displacement convexity of the entropy exactly corresponds to the Prekopa inequality, which expresses more or less the same fact, but at the level of convex functions.

Structure of the paper After this introduction, Section 2 presents the main well-known tools from optimal transport theory (Wasserstein distances, geodesic interpolation...). The estimates and semicontinuity of the entropy term that we provide at the end of the section are also well-known, but presented in alternative fashion. Section 3 is devoted to the transport cost $\rho \mapsto \mathcal{T}(\rho, \mu)$ and to its estimates and semicontinuity in a $\mathcal{P}_1(\mathbb{R}^d)$ framework. Section 4 is the core of the paper, and presents the variational problem that we need to solve in order to find the log-concave measure $\rho = e^{-u}$ that we aim at. In particular, we prove that u is convex and essentially continuous. In Section 5 we show that the condition $(\nabla u)_\# e^{-u} = \mu$ is actually equivalent to the fact that e^{-u} minimizes our functional, using the notion of displacement convexity. Finally, in Section 6 we compare our approach to the one in [9], explaining why they are equivalent and how to pass from one to the other.

Acknowledgments The authors would like to thank Guillaume Carlier and Dario Cordero-Erausquin for interesting discussions about this problem. These discussions have been made possible by the workshop “New Trends in Optimal Transport” organized at the Hausdorff Center for Mathematics in March 2015.

2. FEW WORDS ON OPTIMAL TRANSPORT, ENTROPY AND TECHNICAL TOOLS

We recall here the main notions and notations that we will use throughout the paper. We refer to [13] (Chapters 1, 5 and 7) and to [1, 16] for more details and complete proofs.

Given two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ we consider the set of transport plans

$$\Pi(\mu, \nu) = \{ \gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : (\pi_x)_\# \gamma = \mu, (\pi_y)_\# \gamma = \nu, \}$$

i.e. those probability measures on the product space having μ and ν as marginal measures.

For a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$ we consider the minimization problem

$$\min \left\{ \int c d\gamma : \gamma \in \Pi(\mu, \nu) \right\},$$

which is called the Kantorovich optimal transport problem for the cost c from μ to ν . In particular, we consider the case $c(x, y) = \frac{1}{2}|x - y|^2$. In this case the above minimal value is finite whenever $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, where $\mathcal{P}_2(\mathbb{R}^d) := \{ \rho \in \mathcal{P}(\mathbb{R}^d) : \int |x|^2 d\rho(x) < +\infty \}$.

For the above problem one can prove that the minimal value also equals the maximal value of a dual problem

$$\max \left\{ \int \phi \, d\mu + \int \psi \, d\nu : \phi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2 \right\},$$

and that the optimal function ϕ may be used to construct an optimizer γ . Indeed, the optimal ϕ is locally lipschitz and semiconcave (in particular $x \mapsto \frac{1}{2}|x|^2 - \phi(x)$ is convex) on $\text{spt}(\mu)$ and differentiable μ -a.e. if $\mu \ll \mathcal{L}^d$; one can define a map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ through $T(x) = x - \nabla \phi(x)$ and this map satisfies $T_{\#}\mu = \nu$ and $\gamma_T := (id, T)_{\#}\mu$ (i.e. the image measure of μ through the map $x \mapsto (x, T(x))$) belongs to $\Pi(\mu, \nu)$ and is optimal in the above problem. Moreover, the map T is the gradient of the convex function u given by $u(x) = \frac{1}{2}|x|^2 - \phi(x)$ and is called the optimal transport map (for the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$) from μ to ν . The fact that the optimal transport map T exists, is unique, and is the gradient of a convex function is known as Brenier Theorem (see [5]).

The same could be obtained if one withdrew from the cost $\frac{1}{2}|x - y|^2$ the parts $\frac{1}{2}|x|^2$ and $\frac{1}{2}|y|^2$ which only depend on one variable each (hence, their integral against γ only depends on its marginals). In this case we are interested in a transport maximization problem

$$\max \left\{ \int (x \cdot y) \, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}$$

and the dual problem would become

$$\min \left\{ \int u \, d\mu + \int v \, d\nu : u(x) + v(y) \geq x \cdot y \right\}.$$

In this problem it is quite clear that any pair (u, v) can be replaced with (u, u^*) where $u^*(y) := \sup_x x \cdot y - u(x)$ is the Legendre transform of u and is the smallest function compatible with u in the constraint $u(x) + v(y) \geq x \cdot y$. Then, it is easy to see by the primal-dual optimality conditions that the optimal γ and the optimal u satisfy

$$\text{spt}(\gamma) \subset \{(x, y) : u(x) + u^*(y) = x \cdot y\} = \{(x, y) : y \in \partial u(x)\},$$

which shows that γ is concentrated on the graph of a map T (which is one-valued μ -a.e., provided $\mu \ll \mathcal{L}^d$), given by $T = \nabla u$.

The value of the minimization problem with the quadratic cost may also be used to define a quantity, called Wasserstein distance, over $\mathcal{P}_2(\mathbb{R}^d)$:

$$W_2(\mu, \nu) := \sqrt{\min \left\{ \int |x - y|^2 \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\}}.$$

This quantity may be proven to be a distance over $\mathcal{P}_2(\mathbb{R}^d)$. Moreover, when restricted to the probabilities supported on a given compact set, i.e. to $\mathcal{P}(K)$ with $K \subset \mathbb{R}^d$ compact, it metrizes the weak-* convergence of probability measures. The space $\mathcal{P}_2(\mathbb{R}^d)$ endowed with the distance W_2 is called Wasserstein space of order 2 and denoted in this paper by $\mathbb{W}_2(\mathbb{R}^d)$.

The geodesics in this space play an important role in the theory of optimal transport. If $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu \ll \mathcal{L}^d$, we define $\rho_t := ((1 - t)id + tT)_{\#}\mu$, where T is the optimal transport from μ to ν . This curve ρ_t happens to be a constant speed geodesic for the distance W_2 connecting μ to ν (in case neither μ nor ν are absolutely continuous, it is possible to produce a

geodesic by taking $\rho_t := (\pi_t)_\# \gamma$ where $\pi_t(x, y) = (1-t)x + ty$ and γ is optimal in the Kantorovich problem, which gives the same result if $\gamma = \gamma_T$.

Once we know the geodesics in $\mathbb{W}_2(\mathbb{R}^d)$, one can wonder which functionals $F : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ are geodesically convex, i.e. convex along constant speed geodesics. This notion, applied to the case of the Wasserstein spaces, is also called *displacement convexity* and has been introduced by McCann in [11]. It is very useful both to provide uniqueness results for variational problems and to provide sufficient optimality conditions. A related notion is that of *convexity on generalized geodesics*, which corresponds to $t \mapsto F(\rho_t)$ being convex every time that we take a triplet μ, ρ_0, ρ_1 , the optimal maps T_0 from μ to ρ_0 and T_1 from μ to ρ_1 , and take $\rho_t = ((1-t)T_0 + tT_1)_\# \mu$. The curve ρ_t is not in general the geodesic connecting ρ_0 to ρ_1 . However, the conditions to guarantee displacement convexity and convexity on generalized geodesics are often very similar and most of the useful functionals which are actually used satisfy both notions, and we present this notion here only for the sake of completeness.

We are in particular interested in the entropy functional \mathcal{E} , defined as follows

$$\mathcal{E}(\rho) := \begin{cases} \int \rho(x) \ln(\rho(x)) \, dx & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{otherwise.} \end{cases}$$

This functional is displacement convex as it satisfies the assumption required in [11] (it is also convex on generalized geodesics, but we will not use it here). We also compute its derivative along a geodesic ρ_t . Suppose $\rho_t = (T_t)_\# \rho$, where $T_t = (1-t)\text{id} + tT$. From a simple change of variable, we have

$$\rho_t(T_t(x)) = \frac{\rho(x)}{\det((1-t)I + tDT(x))}$$

(this formula, where DT is the Jacobian matrix of T , is valid provided T_t is countably Lipschitz and injective, which is the case whenever T is an optimal transport). Hence we have

$$\mathcal{E}(\rho_t) = \int \ln(\rho_t(x)) \, d\rho_t(x) = \int \ln \rho_t(T_t(x)) \, d\rho(x) = \mathcal{E}(\rho) - \int \ln(\det((1-t)I + tDT(x))) \, d\rho(x).$$

Hence, we have

$$\frac{d}{dt} [\mathcal{E}(\rho_t)]|_{t=0} = - \int \nabla \cdot (T - \text{id}) \, d\rho.$$

Here the divergence is to be understood in the a.e. sense, as T is countably Lipschitz. If we use $T = \nabla v$ with v convex, this divergence is equal to $\Delta^{ac} v - d$, where Δ^{ac} is the absolutely continuous part of the distributional Laplacian of v , which is a measure.

We also need to underline two other properties of \mathcal{E} , in particular lower bounds and semicontinuity. We stress that both these conditions are easy when we look at measures in $\mathcal{P}(K)$ with $|K| < \infty$, but become trickier on the whole space, because $\rho \ln \rho$ is not positive.

The last property that we need to recall is the semicontinuity of \mathcal{E} w.r.t. the weak-* convergence of probability measures, under the extra condition of a bound on the first moment $\int |x| \, d\rho(x)$. The semicontinuity of functionals of the form

$$\rho \mapsto \int f\left(\frac{d\rho}{d\lambda}(x)\right) \, d\lambda(x)$$

is standard for convex and superlinear f (which is the case for $f(t) = t \ln(t)$) whenever the reference measure λ (which is the Lebesgue measure here) is finite (see for instance [4] or Chapter 7 in [13]). If f is positive it is easy to get the same result by taking the supremum of functionals

restricted to compact subsets, but this is not possible here. Yet, we can notice that for $\rho \ll \mathcal{L}^d$ we have

$$\mathcal{E}(\rho) = \int \left(\rho(x) \ln(\rho(x)) + e^{h(x)-1} - \rho(x)h(x) \right) dx + \int \rho(x)h(x) dx - \int e^{h(x)-1} dx$$

for any function h such that $e^{h(x)-1} \in L^1(\mathbb{R}^d)$ and $h \in L^1(\rho)$. If we suppose $\rho \in \mathcal{P}_1(\mathbb{R}^d)$ we can take $h = -\sqrt{|x|}$. Then we can write $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3$, where

$$\begin{aligned} \mathcal{E}_1(\rho) &= \int \left(\rho(x) \ln(\rho(x)) + e^{h(x)-1} - \rho(x)h(x) \right) dx \\ \mathcal{E}_2(\rho) &= \int \rho(x)h(x) dx, \\ \mathcal{E}_3(\rho) &= - \int e^{h(x)-1} dx. \end{aligned}$$

Notice that the integrand in \mathcal{E}_1 is positive (indeed, for every $a \in \mathbb{R}_+$ and $b \in \mathbb{R}$ we have $a \ln(a) + e^{b-1} \geq a \cdot b$). This allows to write $\mathcal{E} \geq \mathcal{E}_2 + \mathcal{E}_3$, and hence we have

$$\mathcal{E}(\rho) \geq - \int \rho(x) \sqrt{|x|} dx - \int e^{-\sqrt{|x|}-1} dx \geq -C - \sqrt{\int |x| d\rho(x)},$$

where we used $\int d\rho(x) = 1$ in the last inequality (Hölder inequality). This gives a bound from below of $\mathcal{E}(\rho)$ in terms of the square root of the first moment of ρ .

Then, we also express \mathcal{E}_1 as a supremum over compact sets, using the positivity of the integrand:

$$\mathcal{E}_1(\rho) = \sup_{K \subset \mathbb{R}^d, K \text{ compact}} \int \left(\rho(x) \ln(\rho(x)) + e^{h(x)-1} - \rho(x)h(x) \right) dx.$$

Now, we observe that for every sequence $\rho_n \rightharpoonup \rho$ with $\int |x| d\rho_n(x) \leq C$ and $\mathcal{E}(\rho_n) \leq C$ we also have weak convergence $\rho_n \rightharpoonup \rho$ in L^1 , and $\int \psi(x) d\rho_n(x) \rightarrow \int \psi(x) d\rho(x)$ for every sublinear function ψ (i.e. satisfying $\lim_{|x| \rightarrow \infty} \psi(x)/|x| = 0$, which is true for $\psi = h$). Hence, the term \mathcal{E}_1 is l.s.c., \mathcal{E}_2 is continuous and \mathcal{E}_3 is constant and finite for this type of convergence. Globally we can resume these facts in the following proposition.

Proposition 2.1. (1) *The functional $\mathcal{E} : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{+\infty\}$ is well-defined and satisfies*

$$\mathcal{E}(\rho) \geq -C - \left(\int |x| d\rho(x) \right)^{1/2}.$$

(2) *For every sequence $\rho_n \rightharpoonup \rho$ with $\int |x| d\rho_n(x) \leq C$ and $\mathcal{E}(\rho_n) \leq C$ we have $\liminf_n \mathcal{E}(\rho_n) \geq \mathcal{E}(\rho)$.*

(3) *When restricted to $\mathcal{P}_2(\mathbb{R}^d)$, the functional \mathcal{E} is geodesically convex in \mathbb{W}_2 and strictly convex on every geodesic $t \mapsto \rho_t = ((1-t)\text{id} + tT)_\# \rho$ where the map T is not a translation.*

(4) *If $\rho_t = ((1-t)\text{id} + tT)_\# \rho$, then the derivative at $t = 0$ of $t \mapsto \mathcal{E}(\rho_t)$ is given by $-\int \nabla \cdot (T - \text{id}) d\rho$.*

3. THE MAXIMAL CORRELATION FUNCTIONAL

For $\mu, \rho \in \mathcal{P}_1(\mathbb{R}^d)$, with $\int y \, d\mu(y) = 0$, we define the following quantity:

$$\begin{aligned} \mathcal{T}(\rho, \mu) &:= \sup \left\{ \int (x \cdot y) \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\} \\ &= \inf \left\{ \int u \, d\rho + \int u^* \, d\mu : u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\} \text{ convex and l.s.c.} \right\}. \end{aligned}$$

We notice that, if $\rho, \mu \in \mathcal{P}_2(\mathbb{R}^d)$, then we also have

$$\mathcal{T}(\rho, \mu) = \frac{1}{2} \int |x|^2 \, d\rho(x) + \frac{1}{2} \int |y|^2 \, d\mu(y) - \frac{1}{2} W_2^2(\rho, \mu).$$

This functional \mathcal{T} is a transport cost, but we also observe that it is the maximal correlation between ρ and μ , in the sense

$$\mathcal{T}(\rho, \mu) = \sup \{ \mathbb{E}[X \cdot Y] : X \sim \rho, Y \sim \mu \}.$$

For this reason, \mathcal{T} will be called *maximal correlation functional*.

We are interestend in the following properties.

- Proposition 3.1.** (1) For every $\rho \in \mathcal{P}_1(\mathbb{R}^d)$, we have $\mathcal{T}(\rho, \mu) \in [0, +\infty]$.
 (2) If ρ and $\tilde{\rho}$ are one obtained from the other by translation, then $\mathcal{T}(\rho, \mu) = \mathcal{T}(\tilde{\rho}, \mu)$.
 (3) If $\int x \, d\rho_n(x) = 0$ and $\rho_n \rightarrow \rho$, then $\liminf_n \mathcal{T}(\rho_n, \mu) \geq \mathcal{T}(\rho, \mu)$.
 (4) For every $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ with $\int y \, d\mu(y) = 0$ there exists a sequence μ_n of compactly supported measures with $\mu_n \rightarrow \mu$ and $\int y \, d\mu(y) = 0$ such that for every $\rho \in \mathcal{P}_1(\mathbb{R}^d)$ we have $\mathcal{T}(\rho, \mu_n) \rightarrow \mathcal{T}(\rho, \mu)$.
 (5) For every $\rho \in \mathcal{P}_1(\mathbb{R}^d)$ there exists a sequence ρ_n of compactly supported probabilities such that $\rho_n \rightarrow \rho$, $\mathcal{E}(\rho_n) \rightarrow \mathcal{E}(\rho)$ and such that for every $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ with $\int y \, d\mu(y) = 0$ we have $\mathcal{T}(\rho_n, \mu) \rightarrow \mathcal{T}(\rho, \mu)$.

Proof. In order to prove (1), just take $\gamma = \rho \otimes \mu \in \Pi(\rho, \mu)$. We get

$$\mathcal{T}(\rho, \mu) \geq \int (x \cdot y) \, d(\rho \otimes \mu) = \left(\int x \, d\rho(x) \right) \cdot \left(\int y \, d\mu(y) \right) = 0.$$

To prove (2), notice that every $\tilde{\gamma} \in \Pi(\tilde{\rho}, \mu)$ can be expressed as the translation of a $\gamma \in \Pi(\rho, \mu)$, in the sense $\int \phi(x, y) \, d\tilde{\gamma}(x, y) = \int \phi(x + v, y) \, d\gamma(x, y)$, where v is the vector translating ρ into $\tilde{\rho}$. Applying this fact to $\phi(x, y) = x \cdot y$, we get $\mathcal{T}(\tilde{\rho}, \mu) = \mathcal{T}(\rho, \mu) + \int v \cdot y \, d\mu(y) = \mathcal{T}(\rho, \mu)$.

We now prove (3). We restrict to the case $\mathcal{T}(\rho_n, \mu) \leq C$ otherwise the statement is straightforward. We first take optimal plans γ_n such that $\mathcal{T}(\rho_n, \mu) = \int (x \cdot y) \, d\gamma_n(x, y)$. From the tightness assumption on ρ_n , we infer that γ_n are also tight. Hence, we can extract a converging subsequence $\gamma_n \rightarrow \gamma \in \Pi(\rho, \mu)$. If we set $\Gamma_n := \text{spt}(\gamma_n)$, we have a sequence of closed sets in $\mathbb{R}^d \times \mathbb{R}^d$. We can extract a further subsequence locally Hausdorff converging to a closed set Γ . From the cyclical monotonicity of each Γ_n , we infer that Γ is also cyclically monotone. Hence, γ is also optimal, since its support is contained in Γ , which implies that it is cyclically monotone, a condition which is sufficient to guarantee optimality. Hence we have $\mathcal{T}(\rho, \mu) = \int (x \cdot y) \, d\gamma(x, y)$.

We just need to prove $\liminf_n \int (x \cdot y) \, d\gamma_n(x, y) \geq \int (x \cdot y) \, d\gamma(x, y)$. If $x \cdot y$ were a bounded continuous function, we would have equality. The problem is that it is not bounded. Yet, we can prove that it is bounded from below on $\bigcup_n \Gamma_n$, which is enough.

Indeed, take $(x, y), (x', y') \in \Gamma_n$. From cyclical monotonicity we can write

$$x \cdot y + x' \cdot y' \geq x \cdot y' + x' \cdot y.$$

If we integrate the above inequality w.r.t. $d\gamma_n(x', y')$ we get

$$x \cdot y + \mathcal{T}(\rho_n, \mu) \geq 0.$$

Indeed, $\int x \cdot y d\gamma_n(x', y') = x \cdot y$, $\int x' \cdot y' d\gamma_n(x', y') = \mathcal{T}(\rho_n, \mu)$, $\int x \cdot y d\gamma_n(x', y') = 0$ and $\int x \cdot y d\gamma_n(x', y') = 0$. This proves $x \cdot y \geq -C$ on $\bigcup_n \Gamma_n$ and allows to prove (3).

To prove (4), we take for instance $\mu_n := \mu \llcorner B(0, n) + \mu(B(0, n)^c) \delta_{v_n}$, where

$$v_n := \frac{1}{\mu(B(0, n)^c)} \int_{B(0, n)^c} y d\mu(y).$$

In this case we have $\int u^* d\tilde{\mu}_n \leq \int u^* d\mu$ for every convex function u^* , which gives $\mathcal{T}(\rho, \mu_n) \leq \mathcal{T}(\rho, \mu)$. Combining this inequality with the semicontinuity result of (3) we get $\mathcal{T}(\rho, \mu_n) \rightarrow \mathcal{T}(\rho, \mu)$.

The same construction does not work for (5), as we want to guarantee convergence of entropies. In this case, if $\mathcal{E}(\rho) < +\infty$, we need to produce a sequence ρ_n of absolutely continuous measures. We can take $\rho_n := (\rho \llcorner B(0, n)) / \mu(B(0, n))$. In this case we can check explicitly that $\mathcal{E}(\rho_n) \rightarrow \mathcal{E}(\rho)$ by dominated convergence. Moreover, for every convex function u we have $\int u d\rho_n \rightarrow \int u d\rho$ (first we subtract a linear part to u , using $\rho \in \mathcal{P}_1(\mathbb{R}^d)$, and then we are reduced to monotone convergence). Hence, along this sequence, the functional $\mathcal{T}(\cdot, \mu)$ is u.s.c. as an infimum of continuous functional. But the semicontinuity result of (3) provides the continuity. \square

We are also interested in the following estimate on \mathcal{T} as a function of ρ in terms of the first moment of ρ . We first define the constant

$$c(\mu) := \frac{1}{2d} \inf \left\{ \int |y \cdot e - \ell| d\mu(y) : e \in \mathbb{S}^{d-1}, \ell \in \mathbb{R} \right\}.$$

Note that the infimum in the definition of $c(\mu)$ is actually a minimum (we minimize a function which is continuous in e and ℓ , coercive w.r.t. ℓ , and e lives in a compact set). For simplicity, we only state the estimate in the case where ρ is absolutely continuous.

Proposition 3.2. *If $\int x d\rho(x) = 0$ and $\rho \ll \mathcal{L}^d$, then we have $\mathcal{T}(\rho, \mu) \geq c(\mu) \int |x| d\rho(x)$. In particular, \mathcal{T} satisfies an inequality of the form $\mathcal{T}(\rho, \mu) \geq c \int |x| d\rho(x)$ for $c > 0$ and for every ρ with $\int x d\rho(x) = 0$, if and only if μ is not supported on a hyperplane.*

Proof. Take ρ such that $\int x d\rho(x) = 0$ and $\rho \ll \mathcal{L}^d$ and select a vector $e \in \mathbb{S}^{d-1}$ such that $e \mapsto \int (x \cdot e)_+ d\rho(x)$ is maximal. Set $A^+ := \{x : x \cdot e > 0\}$ and $A^- := \{x : x \cdot e < 0\}$. By optimality conditions, this implies that $v^+ := \int_{A^+} x d\rho(x)$ is oriented as e . From the barycenter condition on ρ and the fact that it does not charge the hyperplane $\{x : x \cdot e = 0\}$, the vector $v^- := \int_{A^-} x d\rho(x)$ is opposite to v^+ . Set $m^\pm := \rho(A^\pm)$. We have $m^+ + m^- = 1$, again from $\rho(\mathbb{R}^d \setminus (A^+ \cup A^-)) = 0$.

Let ℓ be a value such that $\mu(\{x : x \cdot e > \ell\}) \leq m^+$ and $\mu(\{x : x \cdot e < \ell\}) \leq m^-$ (if μ does not give mass to the hyperplane $\{x \cdot e = \ell\}$ then both inequalities are equalities). Decompose μ into the sum of two measures μ^\pm such that

$$\text{spt} \mu^+ \subset \{x \cdot e \geq \ell\}, \text{spt} \mu^- \subset \{x \cdot e \leq \ell\}, \mu^\pm(\mathbb{R}^d) = m^\pm.$$

Consider γ obtained in the following way: distribute the mass of $\rho \llcorner A^+$ onto that of μ^+ via a tensor product, and do the same from $\rho \llcorner A^-$ onto μ^- , i.e.

$$\gamma = \frac{1}{m^+}(\rho \llcorner A^+) \otimes \mu^+ + \frac{1}{m^-}(\rho \llcorner A^-) \otimes \mu^- \in \Pi(\rho, \mu).$$

We have

$$\int (x \cdot y) d\gamma = v^+ \cdot \frac{1}{m^+} \int y d\mu^+(y) + v^- \cdot \frac{1}{m^-} \int y d\mu^-(y).$$

We use that v^+ is oriented as e , and $v^- = -v^+$, and we have

$$v^+ \cdot \frac{1}{m^+} \int y d\mu^+(y) + v^- \cdot \frac{1}{m^-} \int y d\mu^-(y) = \frac{|v^+|}{m^+} \int y \cdot e d\mu^+(y) - \frac{|v^+|}{m^-} \int y \cdot e d\mu^-(y).$$

Since μ^+/m^+ and μ^-/m^- are both probability measures, we can subtract the same constant ℓ to the two integrals and get

$$\frac{|v^+|}{m^+} \int (y \cdot e - \ell) d\mu^+(y) - \frac{|v^+|}{m^-} \int (y \cdot e - \ell) d\mu^-(y) = |v^+| \int |y \cdot e - \ell| d\left(\frac{\mu^+}{m^+} + \left(\frac{\mu^-}{m^-}\right)\right).$$

Then, we use $m^\pm \leq 1$ in order to estimate this last result from below with $|v^+| \int |y \cdot e - \ell| d\mu$. In order to conclude, we just need to observe that

$$\int |x| d\rho(x) \leq 2d \sup_{e \in \mathbb{S}^{d-1}} \int (x \cdot e)_+ d\rho(x) = 2d|v^+|.$$

The last part of the statement is easy: if μ is not concentrated on a hyperplane, then $c(\mu) > 0$. If μ is concentrated on a hyperplane H , then take an arbitrary measure ρ concentrated on the line L orthogonal to H and passing through the origin. One can choose it so that $\int x d\rho(x) = 0$ and $\int |x| d\rho(x) > 0$, while $\mathcal{T}(\rho, \mu) = 0$. \square

Finally, we also prove displacement convexity of \mathcal{T} as a function of ρ .

Proposition 3.3. *Let $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$ be absolutely continuous measures, and let $\rho_t = ((1-t)\text{id} + tT)_{\#}\rho$ be the unique constant speed geodesic connecting them for the Wasserstein distance W_2 . Then $t \mapsto \mathcal{T}(\rho_t, \mu)$ is convex on $[0, 1]$. Its derivative at $t = 0$ is larger than $\int (T(x) - x) \cdot y d\gamma(x)$ where γ is an optimal transport plan from ρ_0 to μ .*

Proof. Suppose for a while that $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Then, it is well known that $\rho \mapsto -\frac{1}{2}W_2^2(\rho, \mu)$ is a -1 -displacement convex functional (see [1], Theorem 7.3.2). On the other hand, $\rho \mapsto \frac{1}{2} \int |x|^2 d\rho(x)$ is 1-convex, and the sum of the two, which gives $\mathcal{T}(\rho, \mu) - \frac{1}{2} \int |y|^2 d\mu(y)$, is displacement convex.

To obtain the proof for general μ in $\mathcal{P}_1(\mathbb{R}^d)$, one needs to approximate, and part (4) of Proposition 3.1 allows to do so. Hence, if ρ_t is a geodesic for W_2 , the inequality $\mathcal{T}(\rho_t, \mu_n) \leq (1-t)\mathcal{T}(\rho_0, \mu_n) + t\mathcal{T}(\rho_1, \mu_n)$ passes to the limit and implies the displacement convexity of $\rho \mapsto \mathcal{T}(\rho, \mu)$.

For the last part of the statement, we just observe that

$$\mathcal{T}(\rho_t, \mu) \geq \int (x \cdot y) d((T_t \times \text{id})_{\#}\gamma)(x, y) = \int T_t(x) \cdot y d\gamma(x, y).$$

Differentiating this last term we obtain the desired expression. \square

4. A VARIATIONAL PRINCIPLE FOR MOMENT MEASURES

As we sketched in the introduction, we consider the following variational problem. We fix $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ with $\int y \, d\mu(y) = 0$ and not supported on a hyperplane, and we want to solve

$$(P) \quad \min \left\{ \mathcal{E}(\rho) + \mathcal{T}(\rho, \mu) : \rho \in \mathcal{P}_1(\mathbb{R}^d) \right\}.$$

Theorem 4.1. *The problem (P) admits a solution, which is unique up to translation. If ρ is a solution, and $u : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex l.s.c. function such that $\mathcal{T}(\rho, \mu) = \int u \, d\rho + \int u^* \, d\mu$ with $u = +\infty$ on $\{\rho = 0\}$, then $\rho = ce^{-u}$.*

Proof. To prove existence of a solution, we take a minimizing sequence ρ_n . We can suppose that all ρ_n have 0 as their barycenter as translating them does not change the value of the two parts of the functional. We use the inequality

$$\mathcal{E}(\rho) \geq -C - \left(\int |x| \, d\rho(x) \right)^{1/2},$$

and the inequality $\mathcal{T}(\rho, \mu) \geq c \int |x| \, d\rho(x)$ that we proved in Proposition 3.2. This implies that the moment $\int |x| \, d\rho_n(x)$ must be bounded. In particular, this gives tightness of the sequence ρ_n and we assume $\rho_n \rightharpoonup \rho$. Also, we know that the entropy $\mathcal{E}(\rho)$ is l.s.c. for the weak convergence when the first moment is bounded, and the semicontinuity of \mathcal{T} along sequences with $\int x \, d\rho_n(x) = 0$ was proven in Proposition 3.1.

This proves that a minimizer exists.

We first analyze the optimality conditions: if $\bar{\rho}$ is optimal and u is a convex function realizing the minimum in the definition of $\mathcal{T}(\bar{\rho}, \mu)$, then

$$\rho \mapsto \int \rho \ln \rho \, dx + \int u \, d\rho$$

is minimal for $\rho = \bar{\rho}$. By standard convex minimization arguments this implies that $\bar{\rho}$ also minimizes the linearized functional

$$\rho \mapsto \int \rho (\ln \bar{\rho} + 1) \, dx + \int u \, d\rho,$$

which implies that $\bar{\rho}$ is concentrated on the set of points where $\ln \bar{\rho} + 1 + u$ is minimal. This means that $\bar{\rho} > 0$ on every point where $u < +\infty$, and on these points we need to have $\ln \bar{\rho} = C - u$, i.e. $\bar{\rho} = ce^{-u}$. This same formula also holds on $u = +\infty$, since we necessarily have $\bar{\rho} = 0$ on those points.

In particular, the optimal ρ is a log-concave probability density. This implies that all its moments are finite, and we have $\rho \in \mathcal{P}_2(\mathbb{R}^d)$.

As for uniqueness, we suppose to have two minimizers ρ_0, ρ_1 . We know that they must belong to $\mathcal{P}_2(\mathbb{R}^d)$. We use the displacement convexity of the entropy and of the \mathcal{T} term and observe that the entropy is strictly convex on the geodesic ρ_t unless ρ_0 and ρ_1 are obtained one from the other by translation. This gives uniqueness up to translation. \square

Remark 4.1. *We observe that (P) has no solutions if μ is concentrated on a hyperplane. Suppose this hyperplane is $\{x_d = 0\}$ and take $\rho_n = \frac{1}{2^{d_n}} \mathcal{L}^d \llcorner (R_n)$ where $R_n = \{x = (x_1, \dots, x_d) \in \mathbb{R}^d : |x_i| \leq 1 \text{ for } i = 1, \dots, d-1, |x_d| \leq n\}$. In this case we have $\mathcal{E}(\rho_n) = -\ln(2n)$ and $\mathcal{T}(\rho_n, \mu) \leq \sqrt{d} \int |y| \, d\mu(y)$. Hence $\mathcal{E}(\rho_n) + \mathcal{T}(\rho_n, \mu) \rightarrow -\infty$.*

To prove the main result of the paper, we just need to prove that u is essentially continuous. This can be done in the following way. We recall that, given any solution ρ to problem (P), we can choose as a precise representative of ρ the one given by $\rho = ce^{-u}$, with u convex and l.s.c. (hence ρ is log-concave and u.s.c.).

Theorem 4.2. *Let ρ be the precise representative above of a solution of (P). Set $\Omega = \{u < +\infty\}$. Then $\rho = 0$ \mathcal{H}^{d-1} -a.e. on $\partial\Omega$.*

Proof. Suppose on the contrary that $\rho > 0$ on a set of positive \mathcal{H}^{d-1} measure on $\partial\Omega$. Since Ω is a convex set, we can use local coordinates and assume that this set is of the form $A = \{(x_1, x') \in \mathbb{R} \times \mathbb{R}^{d-1} \mid x' \in B, x_1 = h(x')\}$, where $B \subset \mathbb{R}^{d-1}$ has positive Lebesgue measure and h is a convex real-valued function. In the same chart, Ω would be locally expressed as the set of points satisfying $x_1 > h(x')$. Up to reducing the set A (and hence B), we can suppose that ρ takes values in $[a, b]$, for $0 < a < b < +\infty$. We also observe that ρ , as it is log-concave, is locally bounded. Define $A_\varepsilon := \{(x_1, x') \in \mathbb{R} \times \mathbb{R}^{d-1} \mid x' \in B, x_1 \in [h(x'), h(x') + \varepsilon]\}$. Since $\rho(x_1, x')$ is a continuous function of $x_1 \in [h(x'), h(x') + \varepsilon]$ [(because of its log-concavity and of the representative that we chose), we easily get $\int_{A_\varepsilon} \rho(x) dx = O(\varepsilon)$.

Now, we define a new density ρ_ε as a competitor in (P). We define $T : A_\varepsilon \rightarrow \mathbb{R}^d$ by $T(x_1, x') = (x_1 - \varepsilon, x')$ and we set

$$\rho_\varepsilon = \rho \mathbf{L}(A_\varepsilon^c) + \frac{1}{2} \rho \mathbf{L}(A_\varepsilon) + \frac{1}{2} T_\#(\rho \mathbf{L}(A_\varepsilon)).$$

By computing the density of ρ_ε we can check

$$\mathcal{E}(\rho_\varepsilon) = \mathcal{E}(\rho) - \rho(A_\varepsilon) \ln 2.$$

In order to estimate $\mathcal{T}(\rho_\varepsilon, \mu)$ we take the optimal function u (realizing $\mathcal{T}(\rho, \mu) = \int u d\rho + \int u^* d\mu$) and we modify it into a function u_δ defined as follows. First, take a convex, positive and superlinear function $\chi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\int \chi(x) d\mu(x) < +\infty$. Such a function exists because $\mu \in \mathcal{P}_1(\mathbb{R}^d)$. Then we fix $\delta > 0$ and we take $u = (u^* + \delta\chi)^*$. We have

$$u_\delta(x) \leq u(x) \text{ for every } x \in \Omega \quad u_\delta(T(x)) \leq u(x) + \delta\chi^*\left(\frac{\varepsilon}{\delta}\right) \text{ for every } x \in A_\varepsilon.$$

We have

$$\mathcal{T}(\rho_\varepsilon, \mu) \leq \int u_\delta d\rho_\varepsilon + \int (u^* + \delta\chi) d\mu \leq \mathcal{T}(\rho, \mu) + \frac{1}{2} \rho(A_\varepsilon) \delta\chi^*\left(\frac{\varepsilon}{\delta}\right) + \delta \int \chi d\mu.$$

The optimality of ρ compared to ρ_ε provides

$$\rho(A_\varepsilon) \ln 2 \leq \frac{1}{2} \rho(A_\varepsilon) \delta\chi^*\left(\frac{\varepsilon}{\delta}\right) + \delta \int \chi d\mu.$$

Now, use $\rho(A_\varepsilon) \geq c_0 \varepsilon$ and choose $\delta = c\varepsilon$ for a small constant c such that $c \int \chi d\mu < \frac{1}{2} c_0 \ln 2$. This gives

$$\frac{1}{2} \rho(A_\varepsilon) \ln 2 \leq \frac{c\varepsilon}{2} \rho(A_\varepsilon) \chi^*(c^{-1}),$$

which is impossible as $\varepsilon \rightarrow 0$. □

5. SUFFICIENT OPTIMALITY CONDITIONS

To complete the current study, it remains to prove that every log-concave density $\rho = e^{-u}$ such that $(\nabla u)_\# \rho = \mu$ and u is an essentially continuous convex function $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is necessarily a minimizer of $\mathcal{E}(\rho) + \mathcal{T}(\rho, \mu)$. This would explain that the variational principle of the previous section finds exactly all the desired functions u .

As this result is not the main core of the paper, this section will be little more sketchy than the rest, and will use some results from [9]. Anyway, we claim that the main points of the proof are present in the paper.

The main idea, already investigated in [2, 3] for game theory purposes, is the fact that displacement convexity is sufficient to guarantee minimality when necessary conditions are satisfied.

The result is the following.

Proposition 5.1. *Suppose that $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is an essentially continuous convex function, consider $\bar{\rho} = e^{-u}$ and suppose $(\nabla u)_\# \bar{\rho} = \mu$. Then $\bar{\rho} \in \mathcal{P}_1(\mathbb{R}^d)$ and it solves*

$$\min \left\{ \mathcal{E}(\rho) + \mathcal{T}(\rho, \mu) : \rho \in \mathcal{P}_1(\mathbb{R}^d) \right\}.$$

Proof. Let us consider an arbitrary $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ with compact support and the geodesic $\rho_t = ((1-t)\text{id} + tT)_\# \bar{\rho}$, where $T = \nabla v$ is the optimal transport from $\bar{\rho}$ to ρ . From the compact support assumption on ρ , we get that T is bounded. From the displacement convexity of \mathcal{E} and $\mathcal{T}(\cdot, \mu)$, in order to prove $\mathcal{E}(\rho) + \mathcal{T}(\rho, \mu) \geq \mathcal{E}(\bar{\rho}) + \mathcal{T}(\bar{\rho}, \mu)$ it is sufficient to prove

$$\frac{d}{dt} (\mathcal{E}(\rho_t) + \mathcal{T}(\rho_t, \mu))|_{t=0} \geq 0.$$

The computation of the derivative is included in Propositions 2.1 and 3.3 and we have

$$\frac{d}{dt} (\mathcal{E}(\rho_t) + \mathcal{T}(\rho_t, \mu))|_{t=0} \geq - \int (\Delta^{ac} v - d) d\bar{\rho} + \int (\nabla v(x) - x) \cdot \nabla u(x) \bar{\rho}(x) dx.$$

First we use the inequality

$$d \int e^{-u(x)} dx \geq \int x \cdot \nabla u(x) e^{-u(x)} dx$$

valid for essentially continuous convex functions u . This is actually an equality, but anyway the inequality we need is proven in [9]. Hence

$$\frac{d}{dt} (\mathcal{E}(\rho_t) + \mathcal{T}(\rho_t, \mu))|_{t=0} \geq - \int (\Delta^{ac} v) d\bar{\rho} + \int \nabla v(x) \cdot \nabla u(x) \bar{\rho}(x) dx.$$

We then use $\Delta^{ac} v \leq \Delta v$ (as v is convex and its distributional derivative is a positive measure) and we integrate by parts. We first do it on a ball $B(0, R)$:

$$- \int_{B(0, R)} (\Delta^{ac} v) d\bar{\rho} + \int_{B(0, R)} \nabla v(x) \cdot \nabla u(x) \bar{\rho}(x) dx = - \int_{\partial B(0, R)} \nabla v(x) \cdot n \bar{\rho}(x) d\mathcal{H}^{d-1}(x).$$

We want to pass to the limit as $R \rightarrow \infty$. The second integral may be handled using the fact that $\nabla u(x) \bar{\rho}(x) = -\nabla \bar{\rho}(x)$ and $\nabla \bar{\rho} \in L^1$ (this corresponds to $\mu \in \mathcal{P}_1(\mathbb{R}^d)$) and that $T = \nabla v \in L^\infty$. In the first integral, we use $(\Delta v)(B(0, R)) = \int_{\partial B(0, R)} \nabla v(x) \cdot n d\mathcal{H}^{d-1}(x) \leq CR^{d-1}$, together with the exponential decay of $\bar{\rho} = e^{-u}$, so that $\bar{\rho} \in L^1(\Delta v)$. In the right-hand side, we use again the

exponential decay of $\bar{\rho}$ with the polynomial explosion of $\mathcal{H}^{d-1}(\partial B(0, R))$ and the boundedness of ∇v .

Hence, we get

$$\frac{d}{dt}(\mathcal{E}(\rho_t) + \mathcal{T}(\rho_t, \mu))|_{t=0} \geq \lim_{R \rightarrow \infty} - \int_{\partial B(0, R)} \nabla v(x) \cdot n \bar{\rho}(x) d\mathcal{H}^{d-1}(x) = 0.$$

This proves the optimality of $\bar{\rho}$ when compared to compactly supported measures ρ . for a general ρ , we use part (5) of Proposition 3.1, which allows to approximate every $\rho \in \mathcal{P}_1(\mathbb{R}^d)$ with compactly supported measures ρ_n so that $\mathcal{E}(\rho_n) + \mathcal{T}(\rho_n, \mu) \rightarrow \mathcal{E}(\rho) + \mathcal{T}(\rho, \mu)$. \square

6. FROM THIS VARIATIONAL PROBLEM TO THE ONE STUDIED IN [9]

In this last section we want to make some short comments on (P) and connect it to the problem studied in [9].

From the dual formulation of \mathcal{T} , we may re-write our problem as a min-min problem

$$\min \left\{ \int \rho(x) \ln \rho(x) dx + \int u(x) \rho(x) dx + \int u^*(y) d\mu(y) : \rho \in \mathcal{P}_1(\mathbb{R}^d), u \text{ convex} \right\}.$$

The approach of the present paper consists in considering u as a secondary variable: for every ρ we compute the minimum over possible u , which gives rise to the functional $\mathcal{T}(\rho, \mu)$.

A different possible approach could consist in looking at ρ as the secondary variable, i.e. considering for every u the optimal ρ . It is easy to see that one can compute

$$\inf \left\{ \int \rho(x) \ln \rho(x) dx + \int u(x) \rho(x) dx : \rho \geq 0, \int \rho(x) dx = 1 \right\}$$

by a Lagrange multiplier approach, which means that we need to choose ρ so that $\ln \rho(x) + u = \text{const}$. Hence, ρ is proportional to e^{-u} . Let us write $c = \int e^{-u(x)} dx$. We have $\rho = e^{-u}/c$ and we can compute

$$\int \rho(x) \ln \rho(x) dx + \int u(x) \rho(x) dx = \frac{1}{c} \int (e^{-u}(-u - \ln c) + u e^{-u}) dx = -\ln c.$$

This means that we consider the functional $u \mapsto -\ln(\int e^{-u} dx)$, which is a concave functional of u , and we solve

$$\min \left\{ \int u^* d\mu - \ln(\int e^{-u} dx) : u \text{ convex} \right\}.$$

In the above minimization problem, the first term is convex, since $u \mapsto u^*(y)$ is convex, but the second is concave.

Notice that this transformation of a convex-concave minimization problem (i.e., the minimization of the difference of two convex functions) into another convex-concave minimization is what is usually known as *Toland duality* (see [14, 15]). In particular, we also refer to [8] for the applications of this notion to the case of variational problems involving the term $-W_2^2$, and their connections to variational problems under convexity constraints.

What [9] does, is to consider the same problem in terms of u^* instead of u : the first term becomes linear and the second, magically, convex (thanks to a clever application of a quantitative Prekopa inequality, which corresponds, as we said in the introduction, to the displacement convexity of the entropy). One should not be astonished that they obtain a convex problem: convexity in u^* more or less corresponds to the displacement convexity of the functional that

we study here in terms of ρ (more precisely, convexity in ∇u^* corresponds to the convexity on generalized geodesics with base measure μ).

REFERENCES

- [1] L. AMBROSIO, N. GIGLI, G. SAVARÉ, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Math., ETH Zürich, (2005).
- [2] A. BLANCHET, G. CARLIER, Optimal Transport and Cournot-Nash Equilibria, preprint available at <http://arxiv.org/pdf/1206.6571.pdf>.
- [3] A. BLANCHET, P. MOSSAY, F. SANTAMBROGIO, Existence and uniqueness of equilibrium for a spatial model of social interactions, preprint available at <http://cvgmt.sns.it/paper/1998/>, to appear in *Int. Econ. Rev.*
- [4] G. BOUCHITTÉ, G. BUTTAZZO, New lower semicontinuity results for nonconvex functionals defined on measures, *Nonlinear Anal.*, 15 (7), 679–692, 1990.
- [5] Y. BRENIER, Décomposition polaire et réarrangement monotone des champs de vecteurs. (French) *C. R. Acad. Sci. Paris Sér. I Math.* 305 (19), 805–808, 1987.
- [6] G. BUTTAZZO, *Semicontinuity, relaxation, and integral representation in the calculus of variations*, Longman Scientific & Technical, 1989.
- [7] G. BUTTAZZO, F. SANTAMBROGIO, A model for the optimal planning of an urban area, *SIAM J. Math. Anal.* 37 (2), 514–530, 2005.
- [8] G. CARLIER, Remarks on Toland’s duality, convexity constraint and optimal transport, *Pacific J. Optimization* 4 (3) 423–432, 2008.
- [9] D. CORDERO-EAUSQUIN, B. KLARTAG, Moment measures, *J. Funct. Anal.* 268 (12), 3834–3866, 2015.
- [10] G. DE PHILIPPIS, A. R. MÉSZÁROS, F. SANTAMBROGIO, B. VELICHKOV, BV estimates in optimal transportation and applications, preprint available at <http://cvgmt.sns.it/paper/2559/>.
- [11] R. J. MCCANN, A convexity principle for interacting gases. *Adv. Math.* 128 (1) 153–159, 1997.
- [12] F. SANTAMBROGIO, Transport and concentration problems with interaction effects. *J. Global Optim.*, 38 (1), 129–141, 2007.
- [13] F. SANTAMBROGIO, Optimal Transport for Applied Mathematicians. Birkhäuser, NY, to be published in September 2015.
- [14] J. F. TOLAND, Duality in Nonconvex Optimization, *J. Math. Anal. Appl.*, 66 (2), 399–415, 1978.
- [15] J. F. TOLAND, A duality principle for non-convex optimisation and the calculus of variations, *Arch. Rational Mech. Anal.*, 71 (1), 41–61, 1979.
- [16] C. VILLANI, *Optimal Transport. Old and New*. Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, Berlin, 2009.

LABORATOIRE DE MATHÉMATIQUES D’ORSAY, UNIVERSITÉ PARIS-SUD, 91405 ORSAY CEDEX, FRANCE
E-mail address: filippo.santambrogio@math.u-psud.fr